Research Article | Vol. 14, No. 6 / 1 Jun 2023 / Biomedical Optics Express 3072
Biomedical Optics EXPRESS

Raman spectromics method for fast and label-free genotype screening

SHANSHAN ZHU,[1,2,3,4] YANJIAN LI,[5] FENGDI ZHANG,[2] CHANGCHUN XIONG,[6] HAN GAO,[2] YUDONG YAO,[1] WEI QIAN,[1] CHEN DING,[5,8] AND SHUO CHEN[2,7,9]

[1]*Research Institute of Medical and Biological Engineering, Ningbo University, Ningbo 315211, China*
[2]*College of Medicine and Biological Information Engineering, Northeastern University, Shenyang 110169, China*
[3]*Health Science Center, Ningbo University, Ningbo 315211, China*
[4]*Key Laboratory of OptoElectronic Science and Technology for Medicine of Ministry of Education, Fujian Provincial Key Laboratory of Photonics Technology, Fujian Normal University, Fuzhou 350117, China*
[5]*College of Life and Health Sciences, Northeastern University, Shenyang 110169, China*
[6]*College of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China*
[7]*Key Laboratory of Intelligent Computing in Medical Image, Ministry of Education, Shenyang 110169, China*
[8]*dingchen@mail.neu.edu.cn*
[9]*chenshuo@bmie.neu.edu.cn*

Abstract: It is now understood that genes and their various mutations are associated with the onset and progression of diseases. However, routine genetic testing techniques are limited by their high cost, time consumption, susceptibility to contamination, complex operation, and data analysis difficulties, rendering them unsuitable for genotype screening in many cases. Therefore, there is an urgent need to develop a rapid, sensitive, user-friendly, and cost-effective method for genotype screening and analysis. In this study, we propose and investigate a Raman spectroscopic method for achieving fast and label-free genotype screening. The method was validated using spontaneous Raman measurements of wild-type *Cryptococcus neoformans* and its six mutants. An accurate identification of different genotypes was achieved by employing a one-dimensional convolutional neural network (1D-CNN), and significant correlations between metabolic changes and genotypic variations were revealed. Genotype-specific regions of interest were also localized and visualized using a gradient-weighted class activation mapping (Grad-CAM)-based spectral interpretable analysis method. Furthermore, the contribution of each metabolite to the final genotypic decision-making was quantified. The proposed Raman spectroscopic method demonstrated huge potential for fast and label-free genotype screening and analysis of conditioned pathogens.

1. Introduction

Genes, the fundamental hereditary units, encompass DNA or RNA sequences carrying vital genetic information that offspring can inherit. The expression of genetic information governs the phenotypic variability observed in each individual [1]. Genes and their various mutations have shown substantial associations with the initiation and progression of diseases, including cancer [2,3], diabetes [4], and heart diseases [5]. They serve as important indicators for disease prevention [6], early diagnosis [7], and precision treatment [8,9] in clinical settings. Consequently, genotype screening and analysis have played a significant role in enhancing diagnostic accuracy, therapeutic efficiency, and the prediction of treatment response in personalized medicine.

Nowadays, conventional genotype screening and analysis methods mainly include DNA sequencing, polymerase chain reaction (PCR), and fluorescence in situ hybridization (FISH). However, these methods suffer from limitations such as high cost, time consumption, susceptibility

#493524 https://doi.org/10.1364/BOE.493524
Journal © 2023 Received 18 Apr 2023; revised 22 May 2023; accepted 22 May 2023; published 31 May 2023

Biomedical Optics EXPRESS

to contamination, and complex operation, which hinder their routine use as regular examinations for all patients. DNA sequencing allows the examination of the sequence of four bases (adenine, guanine, cytosine, and thymine) in a specific DNA fragment or even the entire DNA sequence, offering high accuracy. Nevertheless, its cost, time consumption, and complexity restrict its clinical applications [10]. PCR is a widely used molecular biology technique for DNA amplification using enzymatic replication [11,12]. It provides high sensitivity and specificity by selectively amplifying a "target" DNA sequence through primer-mediated enzymatic amplification. However, the amplification process is complex and prone to contamination. FISH is a powerful technique that utilizes fluorescent probes to bind to specific parts of chromosomes with a high degree of sequence complementarity [13]. FISH tests are sensitive and offer hybridization specificity, but they are primarily designed to detect the presence or absence of specific genes and require complex labeling procedures [14]. Therefore, there is an urgent need to develop a rapid, sensitive, and label-free genotype screening method suitable for widespread clinical application.

The integrated study of data from radiographical and genomic scales is known as radiogenomics, which aims to explore the relationship between phenotypic features and genomic alterations [15,16]. However, it has been established that metabolic changes caused by diseases often occur earlier than changes in phenotypic features [17,18]. Therefore, metabolic changes are expected to be closely associated with genotypic variations. Raman spectroscopy, a rapid, sensitive, and label-free technique, can capture metabolic changes by measuring the inelastic scattering of photons from vibrating molecules [19–21]. Inspired by the concept of radiogenomics, we propose a fast, non-invasive, and label-free spectromics method based on Raman spectroscopy. This method investigates the relationship between molecular metabolic features and genomic alterations, as illustrated in Fig. 1. By leveraging big data analysis methods similar to radiogenomics, the Raman spectromics method is expected to establish a more direct and comprehensive association between metabolic changes and genotypic variations for genotype screening.

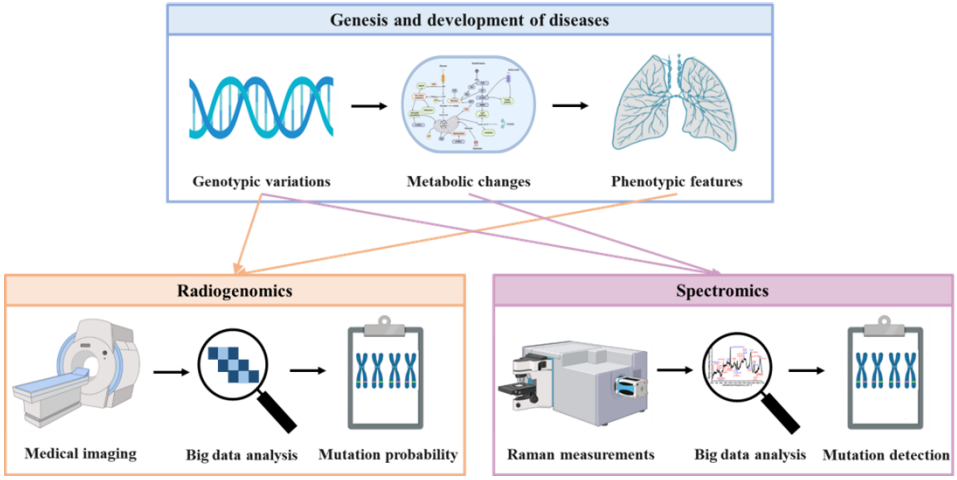

Fig. 1. The schematic of radiogenomics and the proposed spectromics method, in which radiogenomics aims to explore the relationship between phenotypic features and genotypic variations while spectromics aims to create a link between metabolic changes and genotypic variations.

This paper makes the following specific contributions: 1. This study introduces the novel concept of Raman spectromics to examine the correlation between metabolic changes and genotypic variations using a spectroscopic approach. 2. The proposed Raman spectromics method employs a deep learning approach to accurately differentiate between seven strains

Biomedical Optics EXPRESS

of *Cryptococcus neoformans* (*C. neoformans*), providing an alternative method for fast and label-free genotype screening. 3. Using gradient-weighted class activation mapping (Grad-CAM), a hitherto undocumented objective and intelligent spectral interpretable analysis method is introduced to localize genotype-specific regions of interest. Additionally, the contribution of each corresponding metabolite to the final genotypic decision-making is quantified.

2. Materials and methods

The overall framework of the proposed Raman spectromics method is shown in Fig. 2. In this study, the spontaneous Raman spectra were first collected from seven strains of *C. neoformans*; then, a deep learning method was used to classify those spontaneous Raman spectra into different genotypic strains; finally, the contribution of each metabolic change was automatically quantified by interpreting the trained neural network with a newly developed spectral interpretable analysis method.

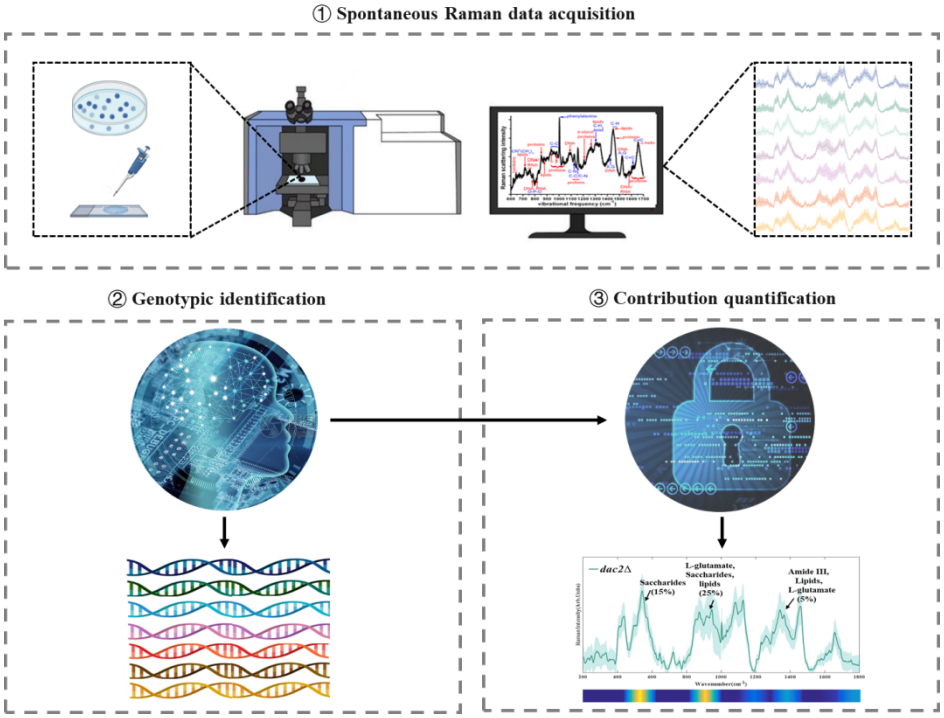

Fig. 2. The overall framework of the proposed Raman spectromics method: ①. Spontaneous Raman data acquisition; ②. Genotypic identification; ③. Contribution quantification.

2.1. Sample preparation and Raman measurements

The flowchart of the sample preparation and Raman measurements is illustrated in Fig. S1 in the Supporting Information. Fungal samples of a wild-type *C. neoformans* strain H99 and six HDAC (histone deacetylase) mutants were cultured in Yeast Extract Peptone Dextrose (YPD) medium. These six HDAC mutants were produced using H99 strain and biolistic transformation, as previously reported [22,23]. According to the closest relative in the phylogenetic tree of deacetylase genes encoding HDACs, the HDAC mutants of *C. neoformans* in this study can mainly be categorized into two classes: class I consisting of *dac2*Δ (Gene ID: CNAG_05563), *dac5*Δ (Gene ID: CNAG_05096), *dac6*Δ (Gene ID: CNAG_05276) and *dac8*Δ (Gene ID:

Biomedical Optics EXPRESS

CNAG_01699) mutants; class II consisting of *dac3*Δ (Gene ID: CNAG_00660) and *dac4*Δ (Gene ID: CNAG_01563) mutants. After overnight cultivation, the fungal samples were diluted in fresh YPD medium and then incubated at 37 °C for 6 h. The mixture was subsequently rinsed and centrifugated 4 times in phosphate-buffered saline (PBS) to remove the culture medium and resuspended in 100 µL PBS solution. Before collecting Raman measurements, a 1 µL suspended fungal sample was dropped into a small well on an aluminum substrate and covered by a quartz coverslip to avoid evaporation.

150 spontaneous Raman spectra were collected from each strain of fungal samples. Thus, 1050 spontaneous Raman spectra were measured from those seven different strains by a confocal Raman microscope (Horiba JY HR Evolution, France). A 785-nm diode laser (about 19.2 mW on the sample) was used as excitation. A laser spot of approximately 1µm was formed by an objective lens (100 x, NA = 0.90, Olympus), which was capable of exciting only a single fungal cell at a time. The measured wavenumber range was from 200 cm^{-1} to 1800cm^{-1} with a spectral resolution of 1 cm^{-1}. The exposure time was 20 s, and each Raman spectrum was accumulated for 10 times.

2.2. Data preprocessing

For each Raman spectrum, the fluorescence background was estimated by fifth-order polynomial fitting and subsequently subtracted from the original spectrum; then, the Savitzky-Golay algorithm was used to remove the noise [21,24]; finally, normalization was implemented to alleviate the variance in Raman measurements due to manual operations, in which the intensity at each wavenumber was divided by the max intensity of this Raman spectrum. After the above operations, those preprocessed Raman spectra served as the inputs for further processing and analysis.

2.3. Genotype screening based on a 1D-CNN model

A 1D-CNN model based on AlexNet architecture [25] was trained to identify fungal samples with different genotypes. The proposed 1D-CNN model consisted of an input layer, four convolutional layers, a flatten layer, three fully-connected layers, and an output layer. The architecture of the 1D-CNN model and its detailed configurations can be found in Fig. 3(a). The preprocessed Raman spectra with each dimension of 2318×1 served as the inputs of the proposed 1D-CNN model. A total of four convolutional layers with different kernel sizes (K), number of filters (F), and strides (S) were employed to extract spectral features. Each convolutional layer was followed by one activation layer and one pooling layer. In the activation layer, a nonlinear activation function ReLU was used to speed up the forward propagation for faster learning and better performance during the training of deep networks [26]. The output of ReLU was the same as the input or zero if the input was positive or non-positive, respectively. In the pooling layer, the max-pooling process was conducted by applying a max filter to non-overlapping subregions of the extracted spectral features. Accordingly, the sizes of those spectral features can be significantly compressed to save computation during the training of deep networks. Those spectral features were further converted into one-dimensional feature vectors by a flatten layer and finally output the predicted genotype after passing through three fully-connected layers. The dropout technique was employed to randomly ignore neurons in the first fully-connected layer with a 50% probability to prevent over-fitting. The proposed 1D-CNN model was coded and run under Python 3.8 with PyTorch 1.11.0.

To carry out an unbiased evaluation, 5-fold cross-validation was employed to train the proposed 1D-CNN model and to validate its performance. The Raman spectra collected from different fungal samples were randomly split into 5 sets with 210 Raman spectra in each set, i.e., 30 Raman spectra randomly chosen from each strain; one set was served as the test data set, and the remaining four sets were further split into training data set and validation data set. Due to the

Biomedical Optics EXPRESS

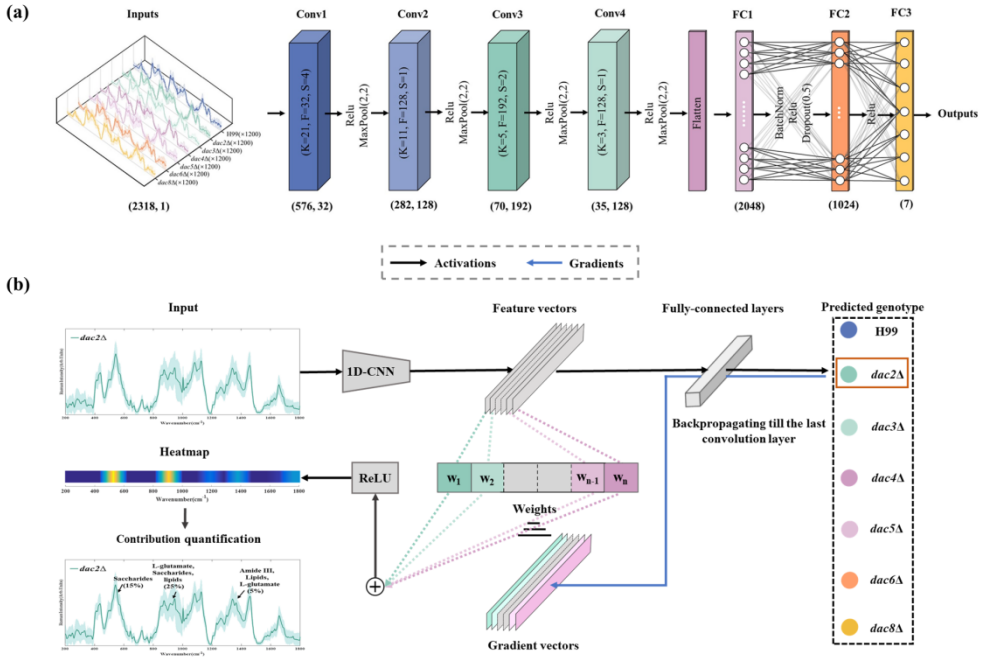

Fig. 3. (a) Architecture of the 1D-CNN model, consisting of an input layer, four convolutional layers, a flatten layer, three fully-connected layers and an output layer. The numbers below each convolutional layer indicate the dimension of the output for this convolutional layer, and the numbers below each fully-connected layer indicate the number of neurons in this fully-connected layer. Conv represents convolutional layer and FC represents fully-connected layer; (b)The schematic of the proposed spectral interpretable analysis method for quantifying the contribution of each metabolic change on final genotypic decision-making. The input Raman spectrum first forward propagated through the trained 1D-CNN model to obtain the final predicted genotype, and then backpropagated to derive the weight of each feature vector; then, the weighted summation of those feature vectors was activated by ReLU to achieve the coarse Grad-CAM localization, in the form of a heatmap along the wavenumber dimension; finally, the contribution of each metabolic change to the final genotypic decision-making is calculated by the division between brightness of corresponding Raman peak and overall brightness of all prominent Raman peaks.

relatively small sample size, data augmentation was applied to training and validation data sets to avoid over-fitting. The sample size was enlarged 10 times by adding white Gaussian noise to the original Raman signal with random noise intensity levels ranging from 0 to 20 dBw. During the training process, considering the compromise between training speed and convergence accuracy, both the training and validation data sets were divided into batches with a batch size of 128. After training, validation, and testing, the performance of the proposed 1D-CNN model was evaluated by the following criteria, including training loss and validation accuracy versus epochs, confusion matrix, and classification accuracy. The training loss indicates the difference between the predicted values of the model and the actual values during the training process, which can be used to adjust model parameters to achieve optimal performance. Validation accuracy refers to the agreement between the predicted results of the model and the actual results. The confusion matrix represents the prediction summary in a matrix form, which shows the number of correct and incorrect predictions in each class. The above procedure was repeated until all 5 sets had been tested once to evaluate the overall performance.

Biomedical Optics EXPRESS

2.4. Grad-CAM based spectral interpretable analysis

This study proposed an intelligent and quantitative spectral interpretable analysis method based on Grad-CAM [27]. This spectral interpretable analysis method could objectively localize the genotype-specific Raman peaks of interest and quantify the contributions of corresponding metabolic changes to the final genotypic decision-making, as demonstrated in Fig. 3(b). More specifically, the input Raman spectrum first forward propagated through the trained 1D-CNN model to obtain the final predicted genotype, and the corresponding gradient vectors generated by the backpropagation of this prediction were global-average-pooled to obtain the weight of each feature vector; then, the weighted summation of those feature vectors was activated by ReLU to achieve the coarse Grad-CAM localization, in the form of a heatmap along the wavenumber dimension, in which the brightness represented the contribution on final genotypic decision-making. Finally, the relative contribution of each metabolic change to the final genotypic decision-making, i.e., C_i was quantified by Eq. (1).

$$C_i = S_i/(S_1 + S_2 + \ldots + S_n) \tag{1}$$

where n is the total number of all prominent Raman peaks, and S_i is the area under full width at half maximum intensity of the *i-th* prominent Raman peak.

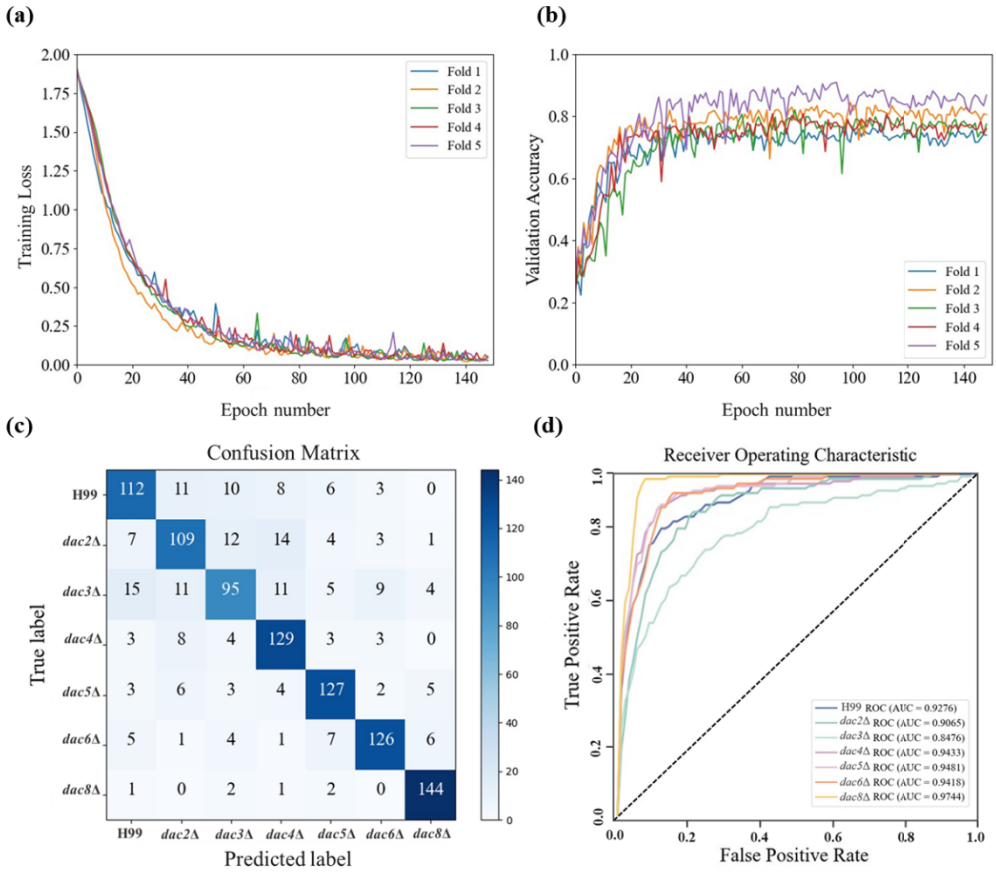

Fig. 4. The training and validation performance versus the epoch number of the proposed 1D-CNN model based on 5-fold cross-validation: (a) training loss, and (b) validation accuracy; The genotype screening performance of the proposed 1D-CNN model on test data set: (c) confusion matrix, and (d) ROC curves.

Biomedical Optics EXPRESS

3. Results

3.1. Genotype screening

The training and validation performances of the proposed 1D-CNN model are depicted in Fig. 4(a) and Fig. 4(b), respectively. The training loss exhibited an initial rapid decrease, followed by a steady decrease until it reached a stable value close to 0, with slight variations. Furthermore, the validation accuracy of the trained 1D-CNN model initially showed a steep increase and then stabilized at around 80% with minor fluctuations. These findings suggest that the trained 1D-CNN model achieved fast convergence without overfitting.

From the confusion matrix in Fig. 4(c), it can be found that most genotypes of *C. neoformans* could be correctly identified. The overall classification accuracy for genotype screening, determined through 5-fold cross-validation, was 80.2%. The 1D-CNN model achieved the highest accuracy of 96% when identifying the *dac*8Δ strain and the lowest accuracy of 63% when identifying the *dac*3Δ strain. These results align with the performances of the 1D-CNN model, as indicated by the ROC curves and AUC values plotted in Fig. 4(d). These findings demonstrate

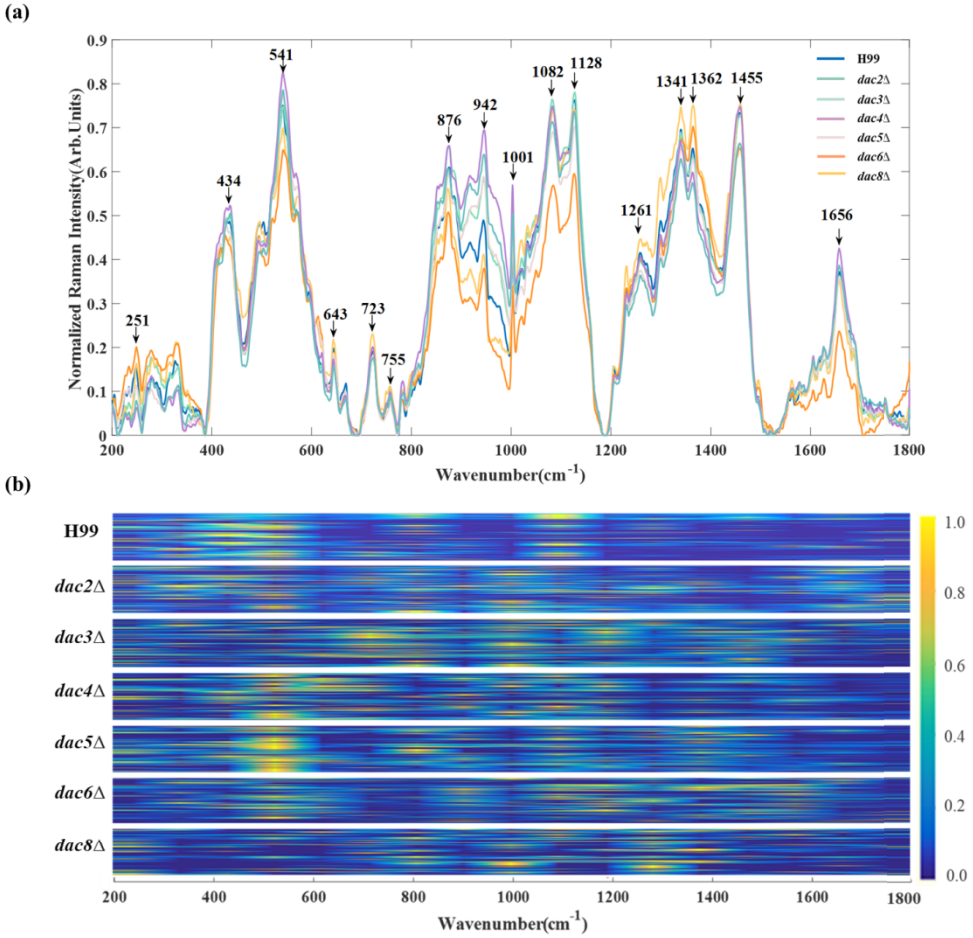

Fig. 5. (a) Comparison of the mean Raman spectra of *C. neoformans* with different genotypes, and (b) the heatmaps, in which the brightness represents the contribution of different Raman peaks to the final genotypic decision-making.

Biomedical Optics EXPRESS

Table 1. The contributions of different metabolic changes to the final genotypic decision-makings for *C. neoformans* with different genotypes

Raman peaks (cm^{-1})	Major assignments	Genotypes						
		H99	*dac2*Δ	*dac3*Δ	*dac4*Δ	*dac5*Δ	*dac6*Δ	*dac8*Δ
251	Chitin	2.3%	1.5%	1.8%	1.5%	1.6%	1.3%	1.2%
434	D-(+)-trehalose	**13.8%**	6.3%	3.7%	6.0%	5.2%	5.0%	3.7%
541	α-D-glucose, β-D-glucose	**29.6%**	**12.1%**	7.4%	**19.6%**	**24.6%**	**17.6%**	8.9%
643	Glutathione	2.1%	1.4%	1.7%	3.4%	2.7%	2.6%	1.2%
723	Adenine, coenzyme A	1.4%	1.7%	3.1%	3.4%	1.0%	1.8%	3.1%
755	L-tryptophan	1.8%	2.1%	4.1%	3.7%	1.0%	1.5%	5.1%
876	L-glutamate	9.3%	9.9%	**10.3%**	4.7%	8.6%	3.8%	**10.6%**
942	L-glutamate, trilinolein, trilinolenin	5.0%	**10.1%**	8.7%	**12.0%**	**11.6%**	8.6%	6.0%
1001	Phenylalanine	0.5%	1.7%	1.7%	1.9%	1.1%	1.4%	1.1%
1082	Amylopectin, amylose	4.4%	**13.0%**	**13.8%**	**11.3%**	4.8%	7.3%	**10.9%**
1128	α-D-glucose, β-D-glucose, D-(+)-trehalose, myristic acid	**11.3%**	7.8%	7.0%	8.9%	6.3%	7.0%	5.7%
1261	D-(+)-trehalose, amide III, trilinolein, trilinolenin	3.6%	6.1%	**11.3%**	8.8%	5.7%	**11.3%**	6.5%
1341	L-glutamate	4.2%	7.4%	7.7%	3.7%	5.6%	9.4%	**11.1%**
1362	Lipids, saccharides	4.1%	6.1%	7.9%	3.8%	9.3%	9.3%	**12.5%**
1455	D-(+)-trehalose, lipids	3.2%	3.8%	7.1%	4.4%	7.6%	7.8%	8.6%
1656	Triolein, trilinolenin, amide I	3.5%	9.1%	2.9%	3.1%	3.4%	4.4%	4.0%

that the combination of spontaneous Raman spectroscopy and deep learning methods enables label-free genotype screening and yields good classification accuracy.

3.2. Genotype spectral interpretable analysis

After preprocessing, the Raman spectra from each genotype of *C. neoformans* are averaged and presented in Fig. 5(a). Upon visual inspection of these mean Raman spectra, consistent and prominent Raman peaks are observed at wavenumbers 251, 434, 541, 643, 723, 755, 876, 942, 1001, 1082, 1128, 1261, 1341, 1362, 1455 and 1656 cm^{-1}. The assignment of these Raman peaks to specific vibrational modes and biomolecules was tentatively done based on existing literature [28–31]. The assigned vibrational modes and corresponding biomolecules are listed in Table S1 in the Supporting Information.

By employing spectral interpretable analysis, the contribution of each metabolic change to the final genotypic decision-making was analyzed and visualized as a heatmap, as shown in Fig. 5(b). The horizontal dimension of the heatmap represents different wavenumbers (i.e., corresponding to different metabolic changes associated with Raman peaks), while the vertical dimension

Biomedical Optics EXPRESS

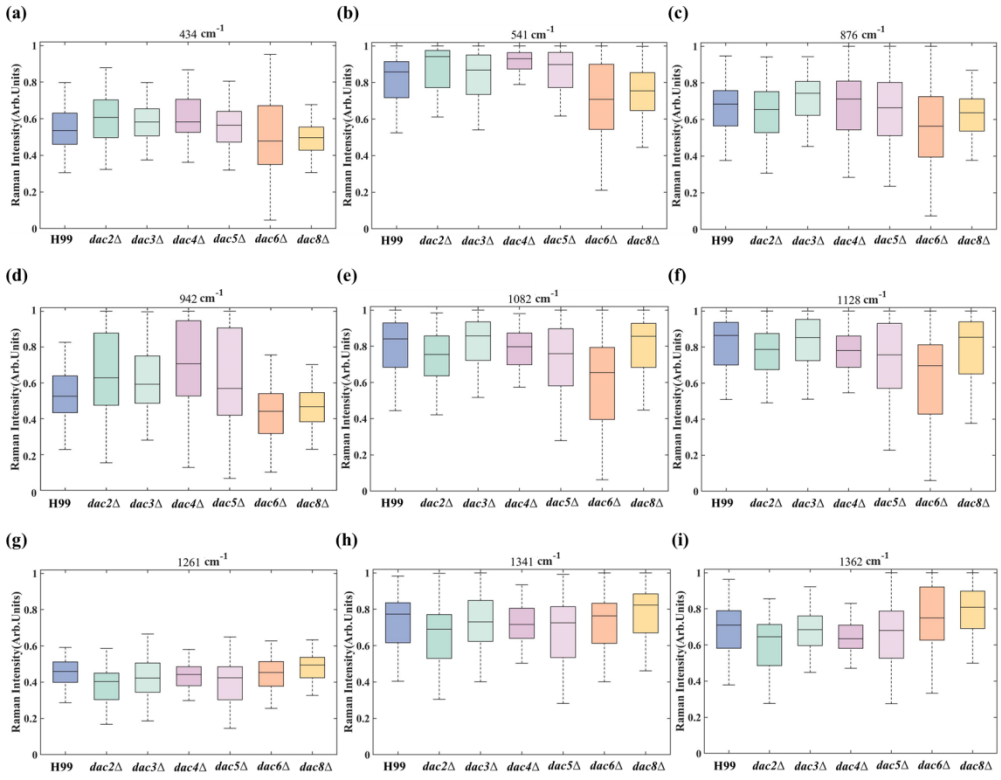

Fig. 6. The Raman intensities of peaks with relative contribution larger than 10% among different genotypes.

represents different fungi samples grouped by genotype. Notably, the heatmap fof each genotype exhibited a strap-like distribution, indicating that the major contributions to the genotypic decision-making originated from consistent metabolic changes within each genotype. However, the metabolic changes with larger contributions varied across different genotypes. Moreover, the relative contribution of each metabolic change was quantified by dividing its individual contribution by the overall contribution from all metabolic changes. The quantified relative contributions are listed in Table 1. Additionally, for Raman peaks with relative contributions exceeding 10%, the Raman intensities across different genotypes are plotted in Fig. 6 to further illustrate the variations.

4. Discussion

It is well-established that genes and their various mutants play a crucial role in the genesis and development of diseases. The field of radiogenomics focuses on the integrated analysis of data from radiographical and genomic scales to investigate the relationship between phenotypic features and genomic alterations. However, it has been shown that metabolic alterations associated with diseases often occur earlier than changes in phenotypic features. Therefore, it is highly conceivable that metabolic changes are more closely linked to genotypic variations. Based on the principles of radiogenomics, we propose a fast, non-invasive, label-free, and cost-effective spectromics method based on Raman spectroscopy to explore the connection between molecular metabolic features and genomic alterations. To validate the effectiveness of this proposed method, we conduct experiments using wild-type *C. neoformans* and its six mutants.

Biomedical Optics EXPRESS

The generation of the six HDAC mutants of *C. neoformans* involved using the H99 strain and biolistic transformation. Histone lysine acetylation/deacetylation is a well-studied histone post-translational modification critical in regulating various cellular processes. It influences chromatin-templated processes and controls metabolic status, cytoskeleton dynamics, apoptosis, protein folding, and cellular signaling in the cytoplasm [32]. This modification is dynamically regulated by the opposing activities of histone acetyltransferases (HATs) and histone deacetylases (HDACs). Several studies have demonstrated that the disruption of the HATs/HDACs balance, resulting from the knockout of specific HDAC genes in fungi, can lead to abnormal gene expression and subsequent changes in fungal phenotypes, growth, metabolism, and virulence [22,23,33]. The HDAC genes in *C. neoformans* have been classified into two classes based on their closest relatives in *S. cerevisiae* or S. pombe: Class I mutants include *dac*2Δ, *dac*5Δ, *dac*6Δ, and *dac*8Δ, while Class II mutants consist of *dac*3Δ and *dac*4Δ [33]. To compare the biomolecular changes between the two classes of mutants and the wild-type H99 strain, the prominent Raman peaks observed in the Raman spectra were assigned to corresponding biomolecules based on Table S1 in the Supporting Information. By analyzing these biomolecular changes, we gained further insights into the metabolic alterations associated with each class of mutants compared to the wild-type strain H99.

In the class I mutants *dac*2Δ and *dac*5Δ, compared to the H99 strain, there was a significant increase in Raman intensity at 942 cm^{-1}, corresponding to L-glutamate. This suggests an elevated level of amino acids in the fungal cells after the deletion of these two histone deacetylase genes. Studies have shown that the deletion of the *DAC2* gene is associated with the upregulation of genes related to protein biosynthesis, indicating its involvement in regulating protein metabolism [34]. Transcriptome analysis also revealed increased glutamine levels upon *DAC2* gene deletion [22]. The increased glutamine is likely converted into L-glutamate by enzymes such as glutaminase, phosphoribosyl pyrophosphate amidotransferase (PPAT), and glutamine-fructose-6-phosphate transaminase [35]. Additionally, it was found that ATPase activity is inhibited after *DAC5* gene deletion [36]. ATPase is involved in various metabolic processes, including protein and lipid metabolism, and its inhibition may contribute to increased amino acid levels. Moreover, the Raman intensities at 1082 cm^{-1} and 1128 cm^{-1} in *dac*2Δ and *dac*5Δ mutants were lower than those in the H99 strain, tentatively indicating changes in saccharides, consistent with the observed capsule defects in *dac*2 Δ and *dac*5Δ mutants, associated with impaired polysaccharide synthases and decreased expression of other cell wall synthesis-related enzymes [34], leading to a reduction in various polysaccharide components of the fungal cell wall and capsule. In contrast, significant changes in other biomolecules were observed in the *dac*6Δ and *dac*8Δ mutants. The Raman intensities at 434 cm^{-1}, 541 cm^{-1}, 876 cm^{-1}, and 942 cm^{-1} in these mutants were lower than those in the H99 strain and were tentatively assigned to D-(+)-trehalose, α-D-glucose, β-D-glucose, L-glutamate, trilinolein, and trilinolenin. Phylogenetic analysis indicated that *Dac*6 and *Dac*8 deacetylases share protein sequence similarities with mammalian histone deacetylases (HDACs) Hdac1 and Hdac2, as well as Rpd3 in *S. cerevisiae* [22,23]. Depletion of Hdac1 has been associated with increased AMP (adenosine 5'-monophosphate) [37]. Interestingly, AMP and ATP can be interconverted to maintain energy balance, and the AMP/ATP ratio is sensed by AMP-activated protein kinase (AMPK), an important regulator of energy homeostasis. Increased AMP levels allosterically activate AMPK, promoting ATP-generating metabolic reactions in various pathways [37]. Thus, the decreased saccharides, lipids, and amino acids observed in *dac*6Δ and *dac*8Δ mutants can be attributed to their massive consumption in metabolism processes such as the tricarboxylic acid cycle, lipid metabolism, and amino acid metabolism. HDAC Rpd3 has also been found to play a crucial role in controlling proliferation, development, and metabolism [38]. In addition, the Rpd3 knockout mutant displays constitutive de-repression of phosphonic acid phosphatase, a key enzyme involved in regulating phospholipid metabolism

Biomedical Optics EXPRESS

and signal transduction by catalyzing the formation of diacylglycerol from phosphonic acid [39], which suggests suppression of lipid levels in *dac*6Δ and *dac*8Δ mutants.

In the class II mutants *dac*3Δ and *dac*4Δ, compared to the H99 strain, there was a significant increase in Raman intensity at 942 cm^{-1}, corresponding to L-glutamate, trilinolein, and trilinolenin. This indicates a decrease in the levels of amino acids and lipids in fungal cells after deleting the *DAC3* or *DAC4* gene. According to the phylogenetic tree analysis of HDAC genes [22,23], HDAC Dac3 shares 35% to 49% homology with Rpd3, while HDAC Dac4 is a homolog of Hda1. These homologous proteins have similar structures, functions, and amino acid sequences. Studies have shown that the deletion of Rpd3 and Hda1 promotes *in vivo* acetylation at various lysine residues in core histones H3 and H4 [40]. Acetylation of highly conserved lysine residues can disrupt more than half of the glycolytic and tricarboxylic acid (TCA) cycle enzymes [41], which may ultimately lead to abnormal amino acid and lipid metabolism. Additionally, Dac4 has been found to regulate a list of histones and other proteins involved in protein synthesis (such as Tef1, mitochondrial nucleoid protein Mnp1, and eukaryotic translation release factor Sup35) [22], which could be another possible mechanism contributing to the decreased amino acid levels in *dac*3Δ and *dac*4Δ mutants. Furthermore, compared to the H99 strain, *dac*3Δ and *dac*4Δ mutants exhibited lower Raman intensity at 1261 cm^{-1}, corresponding to D-(+)-trehalose. This decrease in D-(+)-trehalose levels in *dac*3Δ and *dac*4Δ mutants might be attributed to abnormal glucose metabolism, which the high acetylation levels of lysine residues might also disrupt.

Indeed, the accurate identification of seven different genotypes poses a challenging and complex multi-classification task. While spectral analysis can effectively identify biomolecular differences among groups, it may not be sufficient to solve such a complex classification problem independently. In recent years, deep learning algorithms, such as 1D-CNN, have shown great potential in automated feature extraction and uncovering intricate patterns in high-dimensional data like Raman spectra. By leveraging the power of deep learning, genotype screening among wild-type *C. neoformans* and its mutants can be achieved with high identification accuracy. Furthermore, while estimating the contribution of each biomolecule based on their differences can provide insights, it may not fully capture the significance of these biomolecular changes in the context of classification decision-making. To address this, a Grad-CAM-based method can be employed, which offers a more objective approach to localize genotype-specific Raman peaks of interest for final classification decision-making. This approach allows for more precise and accurate quantification of the contributions of corresponding biomolecular changes to the genotypic decision-making process.

Over the years, gene sequencing, mass spectrometry, and DNA chips have been applied for genotype screening. PCR gene sequencing using specific DNA fragments has demonstrated high sensitivity and specificity in identifying the genotype of *C. neoformans* [42]. However, this method is time-consuming and prone to contamination [43]. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) can classify cryptococci and accurately determine genetic and evolutionary relationships within and between species [44]. Current evidence suggests that MALDI-TOF MS offers higher accuracy and faster results than gene sequencing. However, the cost of the mass spectrometer and its consumables is high, and the outcomes heavily depend on the experience of the technical operator. DNA chips, which are miniaturized microsystems utilizing the specific and reversible binding ability of DNA, have been widely used in mutation detection and identification [45]. Nevertheless, DNA chips are expensive and challenging for routine clinical genotype screening. In contrast, the proposed spectromics method, although validated only on wild-type *C. neoformans* and its six mutations, offers a rapid, sensitive, label-free, and cost-effective approach for genotype screening. Future studies can expand this method to other strains of fungi or bacteria and clinical samples to further demonstrate its broad compatibility and genotype screening capabilities.

Biomedical Optics EXPRESS

5. Conclusions

This study proposed a Raman spectromics method for rapid and label-free genotype screening. The method was assessed using seven *C. neoformans* strains, and a high overall classification accuracy of 80.2% was achieved using a trained 1D-CNN model. A newly developed Grad-CAM-based spectral interpretable analysis method was employed to objectively and intelligently localize genotype-specific Raman peaks of interest. This analysis method also facilitated quantifying the contributions of different metabolic changes to the final genotypic decision-making process. The results suggest significant correlations between phenotype, enzymatic activities, and metabolic processes with genotypic variations in *C. neoformans*. The fast and label-free Raman spectromics method demonstrates the promising potential for genotype screening and analysis of pathogenic organisms.

Funding. National Natural Science Foundation of China (31870140, 32270205, 61605025); Natural Science Foundation of Liaoning Province (2023010756); Natural Science Foundation of Zhejiang Province (LQ23H180004); General scientific Research Project of Zhejiang Education Department Y202146723; Program for the Introduction of High-End Foreign Experts (GXL20200218001); K. C. Wong Magna Fund in Ningbo University.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See Supplement 1 for supporting content.

References

1. P. E. Harrison, A. E. Wright, and J. E. Mank, "The evolution of gene expression and the transcriptome–phenotype relationship," Semin. Cell Dev. Biol. **23**(2), 222–229 (2012).
2. S. R. Morris and L. A. Carey, "Gene expression profiling in breast cancer," Curr. Opin. Oncol. **19**(6), 547–551 (2007).
3. T. Kosaka, Y. Yatabe, H. Endoh, H. Kuwano, T. Takahashi, and T. Mitsudomi, "Mutations of the epidermal growth factor receptor gene in lung cancer: biological and clinical implications," Cancer Res. **64**(24), 8919–8923 (2004).
4. M. Liu, I. Hodish, L. Haataja, R. Lara-Lemus, G. Rajpal, J. Wright, and P. Arvan, "Proinsulin misfolding and diabetes: mutant INS gene-induced diabetes of youth," Trends Endocrinol. Metab. **21**(11), 652–659 (2010).
5. J. D. Curb, R. D. Abbott, B. L. Rodriguez, K. Masaki, R. Chen, D. S. Sharp, and A. R. Tall, "A prospective study of HDL-C and cholesteryl ester transfer protein gene mutations and the risk of coronary heart disease in the elderly," J. Lipid Res. **45**(5), 948–953 (2004).
6. L. Bouchard, I. Blancquaert, F. Eisinger, W. D. Foulkes, G. Evans, H. Sobol, and C. Julian-Reynier, "Prevention and genetic testing for breast cancer: variations in medical decisions," Soc. Sci. Med. **58**(6), 1085–1096 (2004).
7. M. Pradhan, I. Hayes, and A. Vincent, "An audit of genetic testing in diagnosis of inherited retinal disorders: a prerequisite for gene-specific intervention," Clin. Exp. Ophthalmol. **37**(7), 703–711 (2009).
8. K. Rapti, A. H. Chaanine, and R. J. Hajjar, "Targeted gene therapy for the treatment of heart failure," Can. J. Cardiol. **27**(3), 265–283 (2011).
9. W. Burke, D. Atkins, M. Gwinn, A. Guttmacher, J. Haddow, J. Lau, G. Palomaki, N. Press, C. S. Richards, L. Wideroff, and G. L. Wiesner, "Genetic test evaluation: information needs of clinicians, policy makers, and the public," Am. J. Epidemiol. **156**(4), 311–318 (2002).
10. E. Pettersson, J. Lundeberg, and A. Ahmadian, "Generations of sequencing technologies," Genomics **93**(2), 105–111 (2009).
11. J. M. Bartlett and D. Stirling, "A short history of the polymerase chain reaction," Methods Mol. Biol. **226**, 3–6 (2003).
12. A. Wach, A. Brachat, R. Pöhlmann, and P. Philippsen, "New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae," Yeast **10**(13), 1793–1808 (1994).
13. S. M. Wu, X. Zhao, Z. L. Zhang, H. Y. Xie, Z. Q. Tian, J. Peng, Z. X. Lu, D. W. Pang, and Z. X. Xie, "Quantum-dot-labeled DNA probes for fluorescence in situ hybridization (FISH) in the microorganism Escherichia coli," ChemPhysChem **7**(5), 1062–1067 (2006).
14. M. S. Janes, B. J. Hanson, D. M. Hill, G. M. Buller, J. Y. Agnew, S. W. Sherwood, W. G. Cox, K. Yamagata, and R. A. Capaldi, "Rapid analysis of mitochondrial DNA depletion by fluorescence in situ hybridization and immunocytochemistry: potential strategies for HIV therapeutic monitoring," J. Histochem. Cytochem. **52**(8), 1011–1018 (2004).
15. R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: Images Are More than Pictures, They Are Data," Radiology **278**(2), 563–577 (2016).
16. P. Lambin, R. T. H. Leijenaar, and T. M. Deist, *et al.*, "Radiomics: the bridge between medical imaging and personalized medicine," Nat. Rev. Clin. Oncol. **14**(12), 749–762 (2017).

Biomedical Optics EXPRESS

17. B. G. Barwick, M. Abramovitz, M. Kodani, C. S. Moreno, R. Nam, W. Tang, M. Bouzyk, A. Seth, and B. Leyland-Jones, "Prostate cancer genes associated with TMPRSS2-ERG gene fusion and prognostic of biochemical recurrence in multiple cohorts," Br. J. Cancer **102**(3), 570–576 (2010).
18. P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. G. van Stiphout, P. Granton, C. M. Zegers, R. Gillies, R. Boellard, A. Dekker, and H. J. Aerts, "Radiomics: extracting more information from medical images using advanced feature analysis," Eur. J. Cancer **48**(4), 441–446 (2012).
19. A. Downes and A. Elfick, "Raman spectroscopy and related techniques in biomedicine," Sensors **10**(3), 1871–1889 (2010).
20. S. Zhu, Y. Li, H. Gao, G. Hou, X. Cui, S. Chen, and C. Ding, "Identification and assessment of pulmonary Cryptococcus neoformans infection by blood serum surface-enhanced Raman spectroscopy," Spectrochim. Acta, Part A **260**, 119978 (2021).
21. S. Chen, C. Wang, R. Zhu, S. Zhu, and G. Zhang, "Predicting prognosis in acute myeloid leukemia patients by surface-enhanced Raman spectroscopy," Nanomedicine **16**(21), 1873–1885 (2021).
22. Y. Li, H. Li, and M. Sui, *et al.*, "Fungal acetylome comparative analysis identifies an essential role of acetylation in human fungal pathogen virulence," Commun. Biol. **2**(1), 154 (2019).
23. F. Brandão, S. K. Esher, K. S. Ost, K. Pianalto, C. B. Nichols, L. Fernandes, A. L. Bocca, M. J. Poças-Fonseca, and J. A. Alspaugh, "HDAC genes play distinct and redundant roles in Cryptococcus neoformans virulence," Sci. Rep. **8**(1), 5209 (2018).
24. S. Zhu, X. Cui, W. Xu, S. Chen, and W. Qian, "Weighted spectral reconstruction method for discrimination of bacterial species with low signal-to-noise ratio Raman measurements," RSC Adv. **9**(17), 9500–9508 (2019).
25. A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks," Commun. ACM **60**(6), 84–90 (2017).
26. V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning*, Haifa, Israel, Jun (2010).
27. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," Int. J. Comput. Vis. **128**(2), 336–359 (2020).
28. K. C. Schuster, I. Reese, E. Urlaub, J. R. Gapes, and B. Lendl, "Multidimensional information on the chemical composition of single bacterial cells by confocal Raman microspectroscopy," Anal. Chem. **72**(22), 5529–5534 (2000).
29. K. De Gussem, P. Vandenabeele, A. Verbeken, and L. Moens, "Raman spectroscopic study of Lactarius spores (Russulales, Fungi)," Spectrochim. Acta, Part A **61**(13-14), 2896–2908 (2005).
30. C. M. Alunni, P. D. Casagrande, S. Caponi, L. Corte, D. Fioretto, and G. Cardinali, "Meso-Raman approach for rapid yeast cells identification," Biophys. Chem. **254**, 106249 (2019).
31. J. D. Gelder, K. D. Gussem, and P. Vandenabeele, "Reference database of Raman spectra of biological molecules," J. Raman Spectrosc. **38**(9), 1133–1147 (2007).
32. X. J. Yang and E. Seto, "The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men," Nat. Rev. Mol. Cell Biol. **9**(3), 206–218 (2008).
33. F. A. Brandão, L. S. Derengowski, P. Albuquerque, A. M. Nicola, I. Silva-Pereira, and M. J. Poças-Fonseca, "Histone deacetylases inhibitors effects on Cryptococcus neoformans major virulence phenotypes," Virulence **6**(6), 618–630 (2015).
34. S. Liu, Q. Wang, N. Liu, H. Luo, C. He, and B. An, "The histone deacetylase HOS2 controls pathogenicity through regulation of melanin biosynthesis and appressorium formation in Colletotrichum gloeosporioides," Phytopathol Res **4**(1), 21 (2022).
35. S. D. Yelamanchi, S. Jayaram, J. K. Thomas, S. Gundimeda, A. A. Khan, A. Singhal, T. S. Keshava Prasad, A. Pandey, B. L. Somani, and H. Gowda, "A pathway map of glutamate metabolism," J. Cell Commun. Signal. **10**(1), 69–75 (2016).
36. S. Li, Z. Yue, and T.U. Tanaka, "Smc3 deacetylation by hos1 facilitates efficient dissolution of sister chromatid cohesion during early anaphase," Mol. Cell **78**(4), 801 (2020).
37. A. Gonneaud, N. Turgeon, F. M. Boisvert, F. Boudreau, and C. Asselin, "Loss of histone deacetylase Hdac1 disrupts metabolic processes in intestinal epithelial cells," FEBS Lett. **589**(19PartB), 2776–2783 (2015).
38. S. A. Henry and J. L. Patton-Vogt, "Genetic regulation of phospholipid metabolism: yeast as a model eukaryote," Prog. Nucleic Acid Res. Mol. Biol. **61**, 133–179 (1998).
39. M. Vidal and R. F. Gaber, "RPD3 encodes a second factor required to achieve maximum positive and negative transcriptional states in Saccharomyces cerevisiae," Mol. Cell. Biol. **11**(12), 6317–6327 (1991).
40. S. E. Rundlett, A. A. Carmen, R. Kobayashi, S. Bavykin, B. M. Turner, and M. Grunstein, "HDA1 and RPD3 are members of distinct yeast histone deacetylase complexes that regulate silencing and transcription," Proc. Natl. Acad. Sci. U. S. A. **93**(25), 14503–14508 (1996).
41. E. S. Nakayasu, M. C. Burnet, H. E. Walukiewicz, C. S. Wilkins, A. K. Shukla, S. Brooks, M. J. Plutz, B. D. Lee, B. Schilling, A. J. Wolfe, S. Müller, J. R. Kirby, C. V. Rao, J. R. Cort, and S. H. Payne, "Ancient Regulatory Role of Lysine Acetylation in Central Metabolism," mBio **8**(6), e01894 (2017).
42. V. Rivera, M. Gaviria, C. Muñoz-Cadavid, L. Cano, and T. Naranjo, "Validation and clinical application of a molecular method for the identification of Cryptococcus neoformans/Cryptococcus gattii complex DNA in human clinical specimens," Braz. J. Infect. Dis. **19**(6), 563–570 (2015).

Biomedical Optics EXPRESS

43. S. Boondireke, M. Mungthin, P. Tan-ariya, P. Boonyongsunchai, T. Naaglor, A. Wattanathum, S. Treewatchareekorn, and S. Leelayoova, "Evaluation of sensitivity of multiplex PCR for detection of Mycobacterium tuberculosis and Pneumocystis jirovecii in clinical samples," J. Clin. Microbiol. **48**(9), 3165–3168 (2010).
44. A. Panda, A. K. Ghosh, B. R. Mirdha, I. Xess, S. Paul, J. C. Samantaray, A. Srinivasan, S. Khalil, N. Rastogi, and Y. Dabas, "MALDI-TOF mass spectrometry for rapid identification of clinical fungal isolates based on ribosomal protein biomarkers," J. Microbiol. Methods **109**, 93–105 (2015).
45. M. Cuzin, "DNA chips: a new tool for genetic analysis and diagnostics," Transfus. Clin. Biol. **8**(3), 291–296 (2010).